In the beginning was the deed: Verbal short-term memory as object-oriented action

Dylan Jones and Bill Macken

School of Psychology
Cardiff University

Dylan M Jones
Cardiff University School of Psychology
Park Place
Cardiff CF10 3AT

JonesDM@Cardiff.ac.uk

Abstract

Our goal is not to present a new theory of verbal short-term memory (vSTM), but to supplant the concept used to explain performance for some 60 years. We view vSTM and its concomitant processes as reifications from observations of performance in vSTM tasks. Millennia of refining, elaborating and utilising symbolic technologies for representing speech has seduced us into viewing verbal behavior as embodying the hallmarks of such symbolic systems, setting it apart from other types of physical material with which we interact. Contrarily, we maintain that verbal material be seen in the same light as other material. The way in which we encounter and manipulate it (e.g., in the microcosm of the vSTM setting) is to be understood with respect to processes that organise material into perceptual objects that may then be apprehended and manipulated by bodily effector systems. We outline how key empirical hallmarks of vSTM yield to this approach.

Short-term memory (STM) is commonly regarded as a foundation of cognition, a processing primitive underpinning a vast array of mental competencies. Archetypally, the tool used for its study involves the reproduction of a short, serially-presented list of verbal items[1]. Despite the simplicity of the task, people struggle to accurately reproduce sequences of just a few items' length. This is taken to indicate that the cognitive system is underpinned by an inherent limit to 'capacity' posed by the verbal STM system. Furthermore, the variety of settings in which performance in such tasks is associated with broader cognitive, social and developmental functions (both typical and atypical), (for overviews see e.g., Baddeley, 2007; Conway, Jarrold, Kane, Miyake & Towse, 2008; Cowan, 2014) means that our conceptualization of the processes and systems manifest within those tasks becomes incorporated, however implicitly, into the fabric of our conceptualization of the broader functions.

Almost universally, explanations of performance within the vSTM setting are couched in terms of classical cognitive metaphors of encoding, volatile storage, and retrieval of abstract representations of the relevant information. As with other areas of cognition, limits to performance result from operations performed on these representations, stripped of their surface character—spoken, written, or indeed signed—the chief limitation to performance stemming from the fate of the abstract kernel. As with the cognitive paradigm generally, the functioning of vSTM is quintessentially the maintenance and manipulation of discrete, static representations of the nominal units of content (e.g., 'phonemes' or other idealized phonological forms), with the contribution made by perceptual and effector systems being, depending on theoretical variant, supplementary, peripheral, or even epiphenomenal.

Our account not only eschews theoretical concepts such as encoding, storage, retrieval, and memory itself, but also seeks to reconceptualize the very nature of what the tasks entail. We argue that vSTM phenomena are properties of an object-oriented action system involving the organization of the environment into perceptual objects that provide control programs for goal-directed action. Material is presented to the participant, the goal being to organize and apprehend it for re-presentation to the experimenter, whereupon it is evaluated in comparison to a categorical coding scheme. The challenges posed by the task derive from the interaction

[1] For example, a Google Scholar search using "digit span" + "capacity" yields over 17,000 hits from the last decade alone.

between the form in which the material is presented, how it becomes perceptually organized, and how it may be apprehended by an effector system co-opted to meet the specified goal. The temporal aspect of the setting is understood not as *memory* but as the temporal extent in which perception and action take place, just as the casting of an object towards an individual who apprehends and returns it is a temporally-extended process.

Facets of object-oriented action: perception, apprehension, and re-presentation

A key function of perceptual systems is their ineluctable propensity to form objects; products of both the energy arriving at the senses and the past experience of the organism that serves to 'interpret' it. This is the critical characteristic of perception, since it is with objects that automotive organisms must interact. The forms that such interactions may take are constrained both by features of the object and by the availability of an effector system that may be co-opted to accomplish that interaction by apprehending (e.g., grasping, holding, manipulating) the object in a goal-directed fashion (Macken, Taylor & Jones, 2015). The goal that applies in the typical vSTM setting is the re-presentation by the participant, after a short interval, of a version of the content that was originally presented. These three facets combine dynamically to interdependently determine performance.

The effector system that can most readily be co-opted for the apprehension of verbal material, usually, is the speech motor system involved in movement of the vocal tract. A critical, defining characteristic of vSTM tasks is that the content is in dis-integrated form, stripped of the usual semantic, syntactic and prosodic structure of familiar, integrated verbal objects like sentences and phrases[2]. Hence, a disjunction arises between that form and the ready apprehension of the material within the effector system to enable re-presentation according to the task requirements. This disjunction, and the participant's attempts to overcome it, give rise to what appear to be volatility and capacity limitation within the metaphoric vSTM system.

It is via the object-oriented nature of perception that modality impacts on vSTM performance. For example, an acoustically broadly homogenous sequence of words emanating from the

[2] Undoubtedly, some approaches to studying vSTM may use materials that, to varying degrees, incorporate more of those features found in natural language, but here too the conceptual framework for interpreting performance with such materials is that inherited from the study of the meaningless sequence.

same spatial location tends to fuse into a single coherent perceptual object (Bregman, 1990). In comparison, the forces of sequence-level object formation are weaker for the visual analogue. A consequence of object formation (regardless of modality from which the object is transduced) is relatively high-fidelity coding of information at the boundary of that object, with a loss of resolution in the interior (e.g., (Katshu & d'Avossa, 2014; Wagemans et al., 2012). So, auditory presentation leads to U-shaped serial position functions, especially in shorter sequences, revealing successful re-presentation of target content at those boundaries, falling off sharply in the interior. Comparatively, weaker object-formation with visual presentation leads to flatter serial position functions revealing the reduced boundary advantage (i.e., the recency effect: Conrad & Hull, 1964). That visual content has not been bound into a single object means superior resolution of item-by-item sequence content, resulting in medial serial position performance better than auditory presentation. Impeding the apprehension of material by requiring task-irrelevant articulatory activity (articulatory suppression) eliminates this visual advantage, since it resides in the process of apprehension, leaving the boundary advantage for auditory presentation intact, since it resides in the process of obligatory object formation (Macken, Taylor, Kozlov, Hughes, & Jones, 2016).

The advantage for auditory over visual presentation found for boundary content is eliminated by the addition of a redundant item preceding and succeeding the target content (e.g., the word ‘go’ added to a digit sequence. Such effects are typically ascribed to the cognitive processes of interference (overwriting, displacement), but in fact arise from the process of object formation. When the redundant content is fused perceptually with the target content, it usurps the privileged boundary position that would otherwise confer an advantage for that content. Manipulations that serve to organize those redundant items into an object other than the target sequence restore the usual advantage, even though the irrelevant material is still present (Nicholls & Jones, 2002).

Incompatibility between the perceptually rendered object and the task requirement occurs when the target sequence forms more than one object. For example, item-by-item alternation between different voices in an auditory sequence forms two objects, one corresponding to each voice. Neither of these objects contains information in a structure required by the task: re-presenting the verbal items in their original order. The cost associated with dis- and re-assembling these perceptual objects into a task-appropriate form leads to the performance decrement (Hughes, Marsh, & Jones, 2009, 2011), rather than, as is more usually argued, the

cost associated with encoding variable acoustic input into abstract phonological form (e.g., Goldinger, Pisoni, & Logan, 1991; Greene, 1991).

A variety of linguistic variables are also construed in terms of their impact on ready apprehension of the verbal content within the effector system. Volatility of memory is classically attributed to decay during the time taken to refresh/rehearse item representations (e.g., Baddeley, Thomson, & Buchanan, 1975), but it is not duration per se that matters; rather, it is the structural complexity—syllabic, lexical, supra-lexical—of the content (e.g., Caplan, Rochon, & Waters, 1992; Murray & Jones, 2002; Service, 1998). More structurally complex material requires more difficult articulatory configurations to apprehend it in a form that affords faithful re-presentation. Complexity is a product of the material, but also of the participant's experience with it, at both local and global scales (e.g., (Gathercole, Frankish, Pickering, & Peaker, 1999; Woodward, Macken, & Jones, 2008). Not only are sequences of more familiar items (e.g., syllables, words) better recalled than less familiar ones, so too are more familiar sequences (e.g., Jones & Macken, 2015). In both cases, familiarity confers fluency, manifested in smoother articulatory transitions between offsets and onsets of the verbal items within the sequence, and in lenition (vowel or consonant reduction) within familiar lexical and syllabic items (Macken, Taylor, & Jones, 2014).

Conventionally, such effects are modelled as more robust or supplementary encoding or retrieval processes deriving from the existence of long-term lexico-phonological representations (e.g., Jefferies, Frankish, & Noble, 2009; Schweickert, 1993), but key interactions between material, modality, and task suggest that this is not tenable (Macken et al., 2014). So, serial *recall* reveals an advantage for words over non-words, but serial *recognition* (judging whether two sequences are the same) does not. However, this equivalence occurs only in *auditory* recognition; words enjoy an advantage in *visual* recognition. Retrieval constraints are equivalent in both cases (duplicate presentation of the items in original or modified order), so the advantage cannot be due to retrieval processes. Rather, the auditory version of the task may be accomplished without the need to apprehend the content within the effector system (there is no need to re-present it to the experimenter), but it can be based on global matching of the two perceptual objects corresponding to the two sequences. Since visual sequences do not fuse to form single coherent objects, pattern matching as a means of accomplishing the task is unavailable, so the material must be apprehended.

Similarity amongst items within the sequence exerts a powerful effect on performance. Contrary to the common view that this is due to 'phonological' similarity (e.g., Conrad & Hull, 1964; Larsen & Baddeley, 2004), there are in fact two distinct effects of similarity—one perceptual and one motoric—manifested differently depending on sequence position, modality of presentation, requirement to engage in articulatory suppression and presence of to-be-ignored events preceding or succeeding a sequence (Jones, Hughes, & Macken, 2006; Jones, Macken, & Nicholls, 2004; Maidment & Macken, 2012). The perceptual manifestation is better thought of as sequence, or object, homogeneity and is therefore most evident for auditory presentation. Object formation is modulated nonmonotonically by item similarity. At high levels of homogeneity, object formation is strong but content discriminability is weak. As the level of heterogeneity increases discrimination also increases but to some limit where object coherence weakens (such as with a list made up of alternating voices). Effects of discriminability are especially influential at object boundaries where perceptual information is more highly resolved than within the interior of the object. Accordingly, as discussed above, for auditory presentation, the effect of item similarity is especially noticeable at the boundaries of the sequence (and for longer sequences, especially the terminal boundary) (Maidment & Macken, 2012).

Further, similarity amongst items increases errors in the apprehension of the material, given that the similar items afford similar articulatory gestures, leading to reduced correspondence between the apprehended and target forms. Articulatory suppression eliminates that effect of similarity residing in apprehension, while leaving the effect due to perceptual homogeneity intact. This perceptual effect may also be eliminated; redundant auditory events (as discussed above) at the beginning and/or end of the target sequence have their effect by fusing with the target sequence, and occupying the perceptually privileged boundary in their stead (Jones et al., 2004; 2006; Maidment & Macken, 2012).

A further impediment to performance arises when task-irrelevant auditory sequences are present (Salame & Baddeley, 1982). The usual account of the impact of task irrelevant material is that the target content is degraded (over-written, displaced) in the presence of irrelevant material (e.g., Farrell & Lewandowsky, 2002; Oberauer & Lewandowsky, 2008), but evidence shows that this cannot be the case. Not only is the same pattern of disruption found with irrelevant verbal and nonverbal sequences, tasks that require retention of all target

content without reference to its order (e.g., reporting which of a set of well-known items, such as digits, is missing from a sequence), are immune to such disruption, which should not be the case if irrelevant material degraded the representation of the relevant (e.g., Jones & Macken, 1993; Macken, Phelps, & Jones, 2009). Similarly, presenting sequences of irrelevant material such that each verbal token is assigned to a separate auditory object containing repetition of a single token also reduces the impact on performance, even though the same irrelevant content is still present in the setting (Jones & Macken, 1995). Since irrelevant sequences, via the processes of obligatory perceptual organization, constitute alternative objects for the control of action to that of the target sequence, they compete with that sequence for access to the effector system. Thus, if the task itself doesn't require serial re-presentation, or if the irrelevant material doesn't form a sequential object, competition is eliminated and disruption disappears (Macken, Taylor & Jones, 2015).

The process of re-presentation involves mapping from the gradient form in which the material was perceived and apprehended onto the discrete, categorical labels corresponding to the operational definition of the task content. Its success is influenced therefore by the extent to which the participant has knowledge of what those labels may be on a given trial, which may also be influenced by the correspondence between the content and the participant's language knowledge. Its success will also be an outcome of how readily and accurately the material was apprehended initially, as well as the fidelity with which that information may addressed, as a function of object formation processes. This aspect of the setting serves to obscure the dynamics and gradience inherent to object-oriented action since, however it is provided, the participant's response is coded in terms of discrete verbal entities. These are then compared to the presented sequence and evaluated accordingly to infer what happened to that content within the metaphoric store. Since this evaluation only involves comparison of two ordered sequences of discrete entities, it can necessarily only reveal two types of deviation: so-called 'item' and 'order' errors.

However, using a static and discrete coding scheme does not mean that the behavior it is used to evaluate is so. A close and telling parallel to the issues raised by this is found in the study of speech errors, where the coding process involves mapping spoken output onto a phonetic transcription within which only discrete item or order errors can occur (Port, 2010; Port & Leary, 2005). However, when such discrete coding schemata are supplemented by real-time measurement of vocal tract movement (Goldstein, Pouplier, Chen, Saltzman, & Byrd, 2007),

errors in speech are seen to involve graded and simultaneous execution of both target and non-target articulatory configurations. Importantly, such action errors in speech are not always coded in the discrete schemata, either because their acoustic consequences go un-discriminated by the transcriber or because they do not map on to a discrete label within that schema. We are unaware of any attempt to date to analyze vSTM performance in this way, but the lineaments of the dynamic speech motor system become apparent in performance as soon as aspects of prosody—stress, timing, grouping, etc.—are analyzed in the vSTM setting (e.g., Taylor, Macken, & Jones, 2015).

Epilogue

Conventionally, the phenomena that comprise the canon of vSTM sit within the cognitive framework. Indeed, the received view of them contributed to the ascendancy of the cognitive paradigm (Baddeley & Hitch, 1974; Baddeley, 2012). It is a mark of the success and penetration of this cognitive revolution that it became natural to regard human functions as more akin to those of a computer than those of an animal. So deeply entrenched are the tenets of cognitivism that it is difficult to escape their clutches; concepts such as 'encoding', 'storage', 'decay', 'interference', 'retrieval' and even 'memory' and ‘forgetting’ present themselves as self-evident phenomena; things that are simply immanent within observed behavior, rather than hypothetical constructs proposed to explain that behavior (Macken et al., 2015). The framework we outline above starts afresh, addressing the canon of vSTM while abandoning the gamut of cognitivist assumptions along with its argot, and instead understands vSTM as a *setting* in which broader organic functions are seen to be at play.

References

Baddeley, A. (2007). *Working memory, thought, and action*(. Oxford: Oxford University Press.

Baddeley, A., Thomson, N., & Buchanan, M. (1975). Word length and structure of short-term-memory. *Journal of Verbal Learning and Verbal Behavior, 14*(6), 575-589. doi:10.1016/S0022-5371(75)80045-4.

Baddeley, A., & Hitch, G (1974). Working memory. In *The Psychology of Learning and Motivation* (Bower, G.A., ed.), pp. 44-79, Academic Press.

Bregman, A.S., (1990). *Auditory scene analysis: The perceptual organisation of sound*. Boston, Mass.: MIT Press.

Caplan, D., Rochon, E., & Waters, G. (1992). Articulatory and phonological determinants of word-length effects in span tasks. *Quarterly Journal of Experimental Psychology Section a-Human Experimental Psychology, 45*(2), 177-192.

Conrad, R., & Hull, A. (1964). Information, acoustic confusion and memory span. *British Journal of Psychology, 55*(4), 429-432.

Conway, A., Jarrold, C., & Kane, M. (Eds.). (2008). *Variation in working memory*. Oxford: Oxford University Press.

Cowan, N. (2014). Working memory underpins cognitive development, learning, and education. *Educational Psychology Review*, *26*, 197-223.

Farrell, S., & Lewandowsky, S. (2002). An endogenous distributed model of ordering in serial recall. *Psychonomic Bulletin & Review, 9*(1), 59-79. doi:10.3758/BF03196257

Gathercole, S., Frankish, C., Pickering, S., & Peaker, S. (1999). Phonotactic influences on short-term memory. *Journal of Experimental Psychology-Learning Memory and Cognition, 25*(1), 84-95. doi:10.1037//0278-7393.25.1.84

Goldinger, S., Pisoni, D., & Logan, J. (1991). On the nature of talker variability effects on recall of spoken word lists. *Journal of Experimental Psychology-Learning Memory and Cognition, 17*(1), 152-162. doi:10.1037//0278-7393.17.1.152

Goldstein, L., Pouplier, M., Chen, L., Saltzman, E., & Byrd, D. (2007). Dynamic action units slip in speech production errors. *Cognition, 103*(3), 386-412. doi:10.1016/j.cognition.2006.05.010

Greene, R. (1991). Serial-recall of 2-voice lists - implications for theories of auditory recency and suffix effects. *Memory & Cognition, 19*(1), 72-78. doi:10.3758/BF03198497

Hughes, R., Marsh, J., & Jones, D. (2009). Perceptual-gestural (mis)mapping in serial short-term memory: The impact of talker variability. *Journal of Experimental Psychology-Learning Memory and Cognition, 35*(6), 1411-1425. doi:10.1037/a0017008

Hughes, R., Marsh, J., & Jones, D. (2011). Role of serial order in the impact of talker variability on short-term memory: Testing a perceptual organization-based account. *Memory & Cognition, 39*(8), 1435-1447. doi:10.3758/s13421-011-0116-x

Jefferies, E., Frankish, C., & Noble, K. (2009). Lexical coherence in short-term memory: Strategic reconstruction or osemantic glueo? *Quarterly Journal of Experimental Psychology, 62*(10), 1967-1982. doi:10.1080/17470210802697672

Jones, D., Hughes, R., & Macken, W. (2006). Perceptual organization masquerading as phonological storage: Further support for a perceptual-gestural view of short-term memory. *Journal of Memory and Language, 54*(2), 265-281. doi:10.1016/j.jml.2005.10.006

Jones, D., & Macken, W. (1993). Irrelevant tones produce an irrelevant speech effect - implications for phonological coding in working memory. *Journal of Experimental Psychology-Learning Memory and Cognition, 19*(2), 369-381.

Jones, D., & Macken, W. (1995). Organizational-factors in the effect of irrelevant speech - the role of spatial location and timing. *Memory & Cognition, 23*(2), 192-200. doi:10.3758/BF03197221

Jones, D., Macken, W., & Nicholls, A. (2004). The phonological store of working memory: Is it phonological and is it a store? *Journal of Experimental Psychology-Learning Memory and Cognition, 30*(3), 656-674. doi:10.1037/0278-7393.30.3.656

Jones, G. & Macken, B. (2015). Questioning short-term memory and its measurement: Why digit span measures long-term associative learning. *Cognition, 144,* 1-13. doi: 10.1016/j.cognition.2015.07.009

Katshu, M. Z. U. H., & d'Avossa, G. (2014). Fine-grained, local maps and coarse, global representations support human spatial working memory. *Plos One, 9*(9). doi:10.1371/journal.pone.0107969

Larsen, J., & Baddeley, A. (2004). The phonological similarity effect in visual and auditory lists with articulatory suppression. *International Journal of Psychology, 39*(5-6), 595-595.

Macken, B., Taylor, J., & Jones, D. (2014). Language and short-term memory: The role of perceptual-motor affordance. *Journal of Experimental Psychology-Learning Memory and Cognition, 40*(5), 1257-1270. doi:10.1037/a0036845

Macken, B., Taylor, J., & Jones, D. (2015). Limitless capacity: A dynamic object-oriented approach to short-term memory. *Frontiers in Psychology, 6:293*. doi: 10.3389/fpsyg.2015.00293

Macken, B., Taylor, J., Kozlov, M., Hughes, R., & Jones, D. (2016). Memory as embodiment: The case of modality and serial short-term memory. *Cognition, 155*, 113-124. doi:10.1016/j.cognition.2016.06.013

Macken, W., Phelps, F., & Jones, D. (2009). What causes auditory distraction? *Psychonomic Bulletin & Review, 16*(1), 139-144. doi:10.3758/PBR.16.1.139

Maidment, D., & Macken, W. (2012). The ineluctable modality of the audible: Perceptual determinants of auditory verbal short-term memory. *Journal of Experimental Psychology-Human Perception and Performance, 38*(4), 989-997. doi:10.1037/a0027884

Murray, A., & Jones, D. (2002). Articulatory complexity at item boundaries in serial recall: The case of welsh and english digit span. *Journal Of Experimental Psychology-Learning Memory & Cognition, 28*(3), 594-598. doi:10.1037//0278-7393.28.3.594

Nicholls, A., & Jones, D. (2002). Capturing the suffix: Cognitive streaming in immediate serial recall. *Journal Of Experimental Psychology-Learning Memory & Cognition, 28*(1), 12-28. doi:10.1037//0278-7393.28.1.12

Oberauer, K., & Lewandowsky, S. (2008). Forgetting in immediate serial recall: Decay, temporal distinctiveness, or interference? *Psychological Review, 115*(3), 544-576. doi:10.1037/0033-295X.115.3.544

Port, R. (2010). Language as a social institution: Why phonemes and words do not live in the brain. *Ecological Psychology, 22*(4), 304-326. doi:10.1080/10407413.2010.517122

Port, R., & Leary, A. (2005). Against formal phonology. *Language, 81*(4), 927-964. doi:10.1353/lan.2005.0195

Salame, P., & Baddeley, A. (1982). Disruption of short-term-memory by unattended speech - implications for the structure of working memory. *Journal of Verbal Learning and Verbal Behavior, 21*(2), 150-164. doi:10.1016/S0022-5371(82)90521-7

Schweickert, R. (1993). A multinomial processing tree model for degradation and redintegration in immediate recall. *Memory & Cognition, 21*(2), 168-175. doi:10.3758/BF03202729

Service, E. (1998). The effect of word length on immediate serial recall depends on phonological complexity, not articulatory duration. *Quarterly Journal of Experimental Psychology Section a-Human Experimental Psychology, 51*(2), 283-304. doi:10.1080/027249898391639

Taylor, J., Macken, B., & Jones, D. (2015). A matter of emphasis: Linguistic stress habits modulate serial recall. *Memory & Cognition, 43*(3), 520-537. doi:10.3758/s13421-014-0466-2

Wagemans, J., Elder, J. H., Kubovy, M., Palmer, S. E., Peterson, M. A., Singh, M., & von der Heydt, R. (2012). A century of gestalt psychology in visual perception: I. Perceptual grouping and figure-ground organization. *Psychological Bulletin, 138*(6), 1172-1217. doi:10.1037/a0029333

Woodward, A., Macken, W., & Jones, D. (2008). Linguistic familiarity in short-term memory: A role for (co-)articulatory fluency? *Journal of Memory and Language, 58*(1), 48-65. doi:10.1016/j.jml.2007.07.002